www.gencove.com

Imputation accuracy of cattle in ultra lowpass whole genome sequencing

Ultra-low pass sequencing quality of low pass while improving cost and throughput.

INTRODUCTION

- The use of genomic prediction in agriculture has been shown to be an effective way to accelerate breeding by affordably providing accurate predictions of genetic merit. Genome wide genotypes of sufficient density and accuracy can substitute for phenotypic data collection in far less time.
- Several sequencing methods have been adopted to provide genotypes for these applications.
- Here we demonstrate an approach to ultra low-pass (ULP) sequence that will op- timize cost, turnaround time flexibility and accuracy in a benchtop format.
- With a few dollars of sequence data it is possible to drive genomic improvement.

METHODS

- We aimed to evaluate the effectiveness of using a high throughput, ultra low coverage genotyping strategy using efficient library preparation, benchtop sequencers and low-pass sequencing combined with imputation.
- In total, libraries from 1,536 cattle samples (16 replicates each of 96 samples) were prepared using seqWell's plexWell Low Pass 384 kit, converted using the Adept Compatibility Kit v1.1, and sequenced simultaneously using an AVITI sequencer on two flow cells. We targeted average output of less than .2x coverage of the cattle genome.
- Data was demultiplexed and run on the Gencove analysis platform, resulting in imputation of over 170 million SNPs and indels.
- We calculated overall and non-reference concordance of each ultra lowpass repli- cate to its corresponding "truth" set which comprised of the merged FASTQs of each sample (roughly 2x coverage).

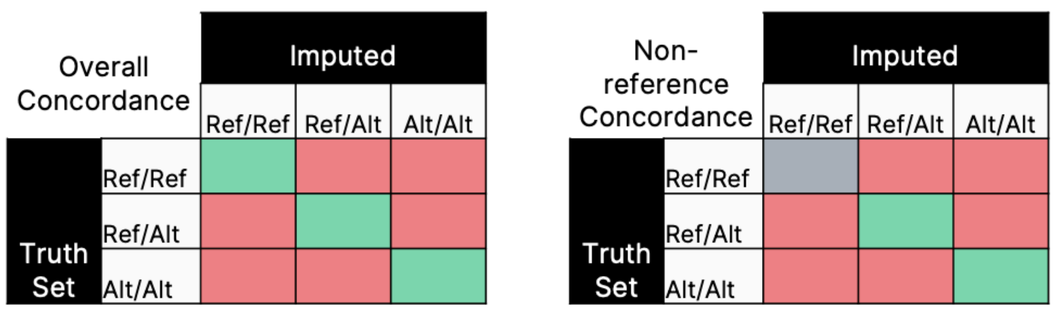

Figure 1: Calculation of overall and non-reference concordance. Metrics are calculated as sum of green cells divided by sum of red cells

DATA

Coverage	0.1X (0.3Gb)	0.15X (0.45Gb)	0.2X (0.6Gb)	0.5X (1.5Gb)
$/sample	$1.5/sample	$2.25/sample	$3/sample	$7.5/sample
Annual Throughput	460k samples	306k samples	230k samples	92k samples

Authors:
Element: Ben Krajacich, Junhua Zhao, Kelly Blease, Semyon Kruglyak
Gencove: Andy Liu, Jahan Parsa, Jesse Hoff
seqWell: Ariele Hanek, Calude Hamby, Kenneth Tenan, Michelle Rahardja

COVERAGE ACROSS REPLICATES

- Raw coverage across all replicates

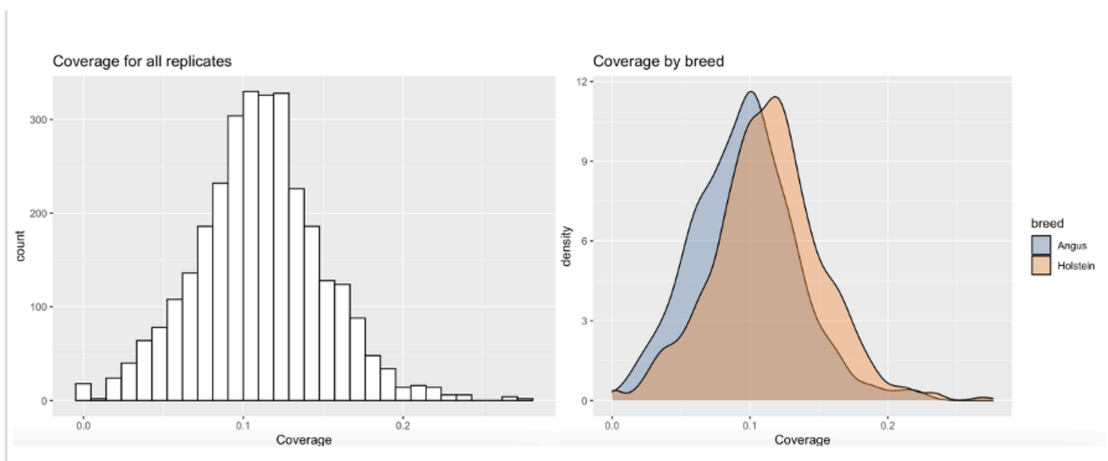

Figure 2: Distribution of coverage for all replicates

CONCORDANCE

- Calculated overall and non-reference concordance for each ultra lowpass replicate with its corresponding higher coverage truth set.

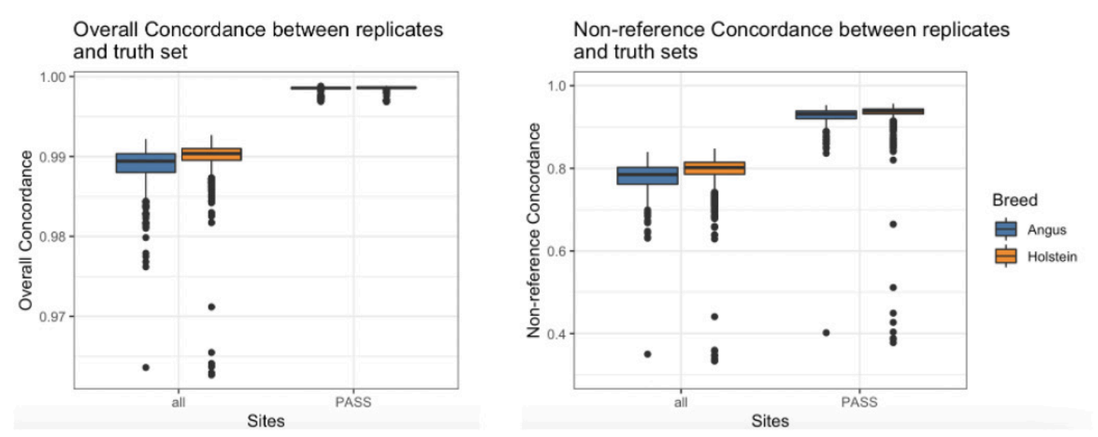

Figure 3: Overall concordance and non-reference concordance

CONCORDANCE AS A FUNCTION OF COVERAGE

- Concordance as a function of coverage

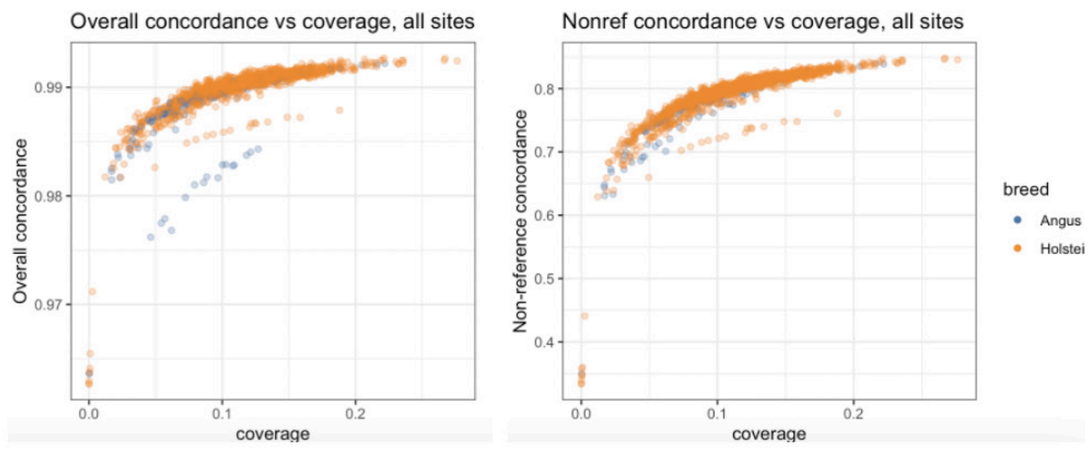

Figure 4: Concordance as a function of coverage

CONCLUSIONS

- At coverages as low as 0.05x, ULP still provides excellent performance giving close to 99% concordance to higher coverage truth sets.
- Small format flow cells can reliably process thousands of samples and deliver consistent levels of coverage to enable an affordable, high throughput genotyping solution, with minimal failure rates.